Facial Behavioral Analysis: A Case Study in Deception Detection

Hoon Yap, Moi; Ugail, Hassan; Zwiggelaar, Reyer

Published in:
British Journal of Applied Science and Technology

DOI:
10.9734/BJAST/2014/6369

Publication date:
2014

Citation for published version (APA):
Hoon Yap, M., Ugail, H., & Zwiggelaar, R. (2014). Facial Behavioral Analysis: A Case Study in Deception Detection. *British Journal of Applied Science and Technology*, *4*(10), 1485-1496.
https://doi.org/10.9734/BJAST/2014/6369

Research Paper

Facial Behavioral Analysis: A Case Study in Deception Detection

ABSTRACT

Aims: To establish a rich Facial Action Coding System (FACS) coded database and to investigate the use of the faces visual cues for deception detection.
Study design: A within-participants design experiment was conducted, using immigration as a scenario for asking questions of participants in control and experimental conditions. The study design required participants to answer questions on two topics, one as themselves and one falsely. Data regarding visible images of facial movement were collected and analyzed against cues identified as indicative of deceit.
Place and Duration of Study: With the ethical approval from University of Bradford, 32 volunteer undergraduate students and research assistants were took part in the study, from March 2011 – June 2011.
Methodology: We included 32 students (27 men, 5 women; age range 18-33 years). The experiment was constructed as two interview scenarios. Participants were interviewed by an 'Examiner' who was introduced by the 'Facilitator' as having recently trained in techniques to detect lie. Participants were told it was important that they appear honest throughout. For one session, they were asked to answer questions as themselves. For the other, they were given a character profile to learn and were asked to answer the questions as if they were the character in the profile. Some questions went beyond the information in the profile, requiring participants to create plausible answers. A rich Facial Action Coding System coded (FACS-coded) database was established for further analysis.
Results: The Examiner's score is 56.25% in both sensitivity and specificity. The best classification algorithm for our FACS-coded database was Logistic Regression with a sensitivity of 47.9%, and a specificity of 71.2%. The findings revealed that the machine learning were bias to truth. In order to increase the sensitivity of deceit prediction, the threshold of classification was adjusted, and the improved result indicates sensitivity of 70.0% and specificity of 63.3%.
Conclusion: Our research established a rich FACS coded database that is important in future research development. In order to increase the detection rate, we proved that it is worthwhile to consider machine learning algorithms to aid human decision.

Keywords: facial behavioral analysis, deception, FACS coding, machine learning, classification.

1. INTRODUCTION

An emerging theme of interest for security agencies is the detection of human behaviors that may reveal an individual as having deliberate malicious intent; for instance by attempting to deceive authorities to enter a country illegally, smuggle goods into or out of a country, being involved in a malicious act such as a terrorist bombing, or as harboring the intention to carry out such a malicious act at a later time. Such a capability will aid in the apprehension of suspect individuals, before they are able to carry out malicious acts. Relevant literatures were reviewed to establish behaviors that might plausibly be used for the operational identification of malicious intent: modeling these behaviors, patterns or cues will provide a significant base for a tool in detecting suspicious individuals.

Most people believe that they can tell when someone is lying to them. However, the evidence from psychology experiments shows that, on average, people only discriminate liars from truth tellers in about 54% of cases [1, 2]. This performance does not represent a very meaningful improvement over chance [3, 4]. However, evidence shows that the performance in deception detection is higher in high-stakes [5]. Researchers [6, 7] do suggest liars behave differently from truth tellers—and so might be identifiable—because the process of lying initiates three psychological constructs: emotion [8, 9]; content complexity [9, 10]; and attempted control [10].

For example, people who are lying might be expected to experience 'emotions' including guilt, fear and duping delight [8]. They will also experience 'content complexity' due to having to 'check their story' to ensure its consistency and believability. This includes thinking of plausible answers to questions, avoiding contradictions, making sure lies are compatible with other available information and remembering what they have said so they can repeat it later and will increase the cognitive workload in comparison to someone telling the truth [9-11]. Liars will also be concerned about behaviors that could give them away, so need to control their actions—described as 'impression management' (Krauss, 1981, cited in Bull et al., 2002 [3]). Research shows that this often creates an over-compensation [3, 10, 12] which might be detectable, and also reinforces the increased cognitive load associated with lying. Indicators that an individual is experiencing any one of these psychological constructs might therefore indicate their attempt to deceive and so identify them for further questioning. Moreover, it is likely that the dominance of each construct over the others will vary through the narrative of a security process. Appreciation of this variation will vastly enhance the effectiveness of any tool used to detect those with malicious intent.

Alongside these three constructs, there are other necessary considerations. Cues related to anxiety, for example, may be more difficult to detect in less trait-anxious individuals [13], or those who are experienced at deception. Furthermore, innocent individuals may display signs of anxiety since emotions are likely to always 'run high' in security settings, for a variety of reasons. The difference between 'state' and 'trait' anxiety therefore becomes pertinent. State anxiety is a temporary feeling of anxiety experienced as a result of an external influence whereas trait anxiety is the individual's general tendency to respond with anxiety to perceived threats: the 'individual differences' between people in terms of their experience of state and trait anxiety will impact on their behavior in security settings. These points suggest that the cues that indicate a high cognitive load or attempts at control may be more promising as operational indicators of deception since they are less likely to appear in innocents.

In terms of emotion expression within the face, some researchers believe there are different elements of specific expressions corresponding with specific emotions [14]. Others argue for a more general dimensionality [15]. Cultural display rules affect the relationship between feeling and display, people can exaggerate or hide expressions to conform to accepted patterns [8], and there are questions about whether emotions can be expected to have basic links to expressions, or whether the face is simply a tool for communicating intentions [16, 17]. Research [18, 19] suggest that rich media or multimodalities provides more clues in term of synchronicity and consistency of the communication. In communication theory, deception principles were merged with interpersonal communication principles [20]. There may be common clues to 'abnormal' behavior, or to attempts to conceal feelings, as they will not always (depending on the skill of the individual) appear the same as natural, unchecked expressions. Therefore a baseline was sought to understand facial behavior in truthful and deceitful scenarios, to enable development of a suitable decision-aid tool.

(a) (b) (c)

Figure 1: Experimental equipment setup: (a) facilitator briefs the participant, (b) interview session, (c) visual camera model.

An experiment was constructed to establish a baseline of the specified behaviors in truthful and deceitful conditions. A rich FACS-coded (Facial Action Coding System) database was established from the baseline data to support future development of a tool for operational detection of cues to malicious intent. The detail description of FACS and the annotation is described in *section 3.2*. The FACS-coded database will also aid the psychology and computer vision communities as there is currently a data deficit in this area.

2. MATERIAL AND METHODS

2.1 Protocol

The experiment was constructed as two interview scenarios. Participants were interviewed by an 'Examiner' who was introduced by the 'Facilitator' as having recently trained in techniques to detect lie. Participants were told it was important that they appear honest throughout. For one session, they were asked to answer questions as themselves. For the other, they were given a character profile to learn and were asked to answer the questions as if they were the character in the profile. Some questions went beyond the information in the profile, requiring participants to create plausible answers.

Each session consisted of a period of introduction followed by a series of five introductory questions (for example, 'what is your name?') asked by the Facilitator, followed by an interview with the second experimenter: the 'Examiner' who asked 10 set questions on the relevant topic. Throughout the experiment, the data regarding visible images of facial behavior were coded by certified FACS coders.

A within subjects approach was employed with two independent variables: interview topic (university study and career, dwelling hobbies personality and family) and honesty (self, character). Condition orders were counterbalanced, as shown in Table 1, and the interviewer was blind to the condition to prevent bias. Participants were invited for two interview sessions, one in the morning and one in the afternoon of the same day. This provided separation between the two topics and the truthful and deceitful conditions.

The questions were designed to elicit answers of 2 to 10 seconds in the majority of questions. It was anticipated that this would be sufficient, combined with measurement of facial behavior during the question period, to represent the range of facial behavior satisfactorily. In the next section, we provide further description about the equipment setup.

2.3 Equipment Setup

The experiment was conducted in a dark room with controlled lighting condition. Figure 1(a) illustrates the facilitation session, while facilitator was giving the instructions to the participant. Figure 1(b) illustrates the position of the environment during the interview session. The participants facial activitities were recorded by using a high definition visual camera, as illustrated in figure 1(c). The model of high definition camera used in this experiment is JVC-GY-HM100E, we set the resolution to 1280 by 720.

2.4 Examiner

During the interview, the examiner dressed formally to reinforce the impression of authority. The examiner was blind in that he did not know about the design of the study or which condition a participant would be in. He was not involved in the day to day running of project. To enable rewards to be given to participants as an incentive, the examiner recorded his judgment as to whether each participant was telling the truth but was not told whether his judgment was correct.

Although not the focus of the experiment, it may be noteworthy that the Examiner who took part in the study is an expert in crime scene reconstruction and forensic science.

2.5 Facilitator

The experiment was fully facilitated using scripted participant introduction and instructions. The facilitator mentioned the ‘examiner’ and informed the participant that the examiner has been trained in techniques for detecting lies. Then, the facilitator explained that the examiner would interview the participant on two topics and informed the participants that the trial is designed to investigate methods for detecting when someone is lying.

Finally, the facilitator reminded the participant of the importance of presenting themselves as honest throughout the entire interview, and, if appropriate of staying consistent and in character for the relevant topic. The participant was informed by the facilitator that there was a small reward available for those participants who convince the examiner that they are truthful throughout the interview.

2.6 Participants

With the ethical approval from University of Bradford, 32 volunteer undergraduate students and research assistants were took part in the study. Among them, 27 were male and 5 were female. They ranged from 18 years to 33 years.

2.7 Self-report

At the end of each session, the participant was asked to confirm whether they had followed the instructions correctly and answered as themselves or the character (as appropriate) for each question. The facilitator also thanked the participant for their participation, informed the participant of the examiner’s judgment and provided a small reward if the participant was successful in convincing the examiner that they were truthful throughout the interview.

2.8 Analysis

Facial behavior was measured throughout the interview sessions, during both the introductory questions, and the interview with the Examiner. Facial indicators are likely to occur throughout listening and preparation of an answer, therefore participant behaviors were analyzed for both question and response periods. The measure of facial behavior was done manually by FACS coders. To avoid bias scores, the FACS coders did not know the condition of the coding or the meaning of the cues.

Table 1. Participant Ordering and Topic Ordering

Subject	First session	Second session
1	Topic A – lie	Topic B - truth
2	Topic A – truth	Topic B – lie
3	Topic B – lie	Topic A – truth
4	Topic B – truth	Topic A – lie
5	Topic A – lie	Topic B - truth
6	Topic A – truth	Topic B – lie
7	Topic B – lie	Topic A – truth
8	Topic B – truth	Topic A – lie
9	Topic A – lie	Topic B - truth
10	Topic A – truth	Topic B – lie
11	Topic B – lie	Topic A – truth
12	Topic B – truth	Topic A – lie
13	Topic A – lie	Topic B - truth
14	Topic A – truth	Topic B – lie
15	Topic B – lie	Topic A – truth
16	Topic B – truth	Topic A – lie
17	Topic A - lie	Topic B - truth
18	Topic A – truth	Topic B – lie
19	Topic B – lie	Topic A – truth
20	Topic B – truth	Topic A – lie
21	Topic A - lie	Topic B - truth
22	Topic A – truth	Topic B – lie
23	Topic B – lie	Topic A – truth
24	Topic B – truth	Topic A – lie
25	Topic A - lie	Topic B - truth
26	Topic A – truth	Topic B – lie
27	Topic B – lie	Topic A – truth
28	Topic B – truth	Topic A – lie
29	Topic A - lie	Topic B - truth
30	Topic A – truth	Topic B – lie
31	Topic B – lie	Topic A – truth
32	Topic B – truth	Topic A – lie

3. RESULTS AND DISCUSSION

We discuss the results from two perspectives: first analysis from human judgment (Examiner's judgment) based on verbal and non-verbal cues; and second is to explain the process in the database preparation and the performance of the computer algorithms based on the FACS-coded database, to aid human decision.

3.1 Analysis on Examiner's score

The examiner's judgment provided a means to incentivize and reward participants; it was not the focus of this research. Research showed that average person spots liars at approximately 54% accuracy [1], while the specialized groups (trained psychologist, police etc.) score approximately 60% accuracy in identifying deception [21].

The confusion matrix of the examiner's score in detecting deception is presented in Table 2, which shows that the Examiner achieved 56.25% accuracy in detecting truth tellers and 56.25% in detecting deceit. The sensitivity and specificity of 56.25% revealed the weakness of human in deception detection. The next section presents discussion of the analysis of visual facial cues as an indicator of deceit.

Table 2. Confusion Matrix on Examiner's Score

	Predicted Class			
		Lie	**Truth**	**Total**
Actual Class	**Lie**	18	14	32
	Truth	14	18	32
	Total	32	32	64

3.2 FACS Coding Annotation

The Facial Actions were coded using FACS [22]. FACS provides comprehensive and objective way to analyze expressions into elementary components. It has been used widely in behavioral sciences. All the action units were coded by certified FACS coders. In our investigation, the duration of an action unit is the total time taken from onset, apex, and offset. Besides the standard AUs, we also analyzed behaviors related to anxiety such as gaze, stuttering, swallowing, and lip biting. For FACS annotation, we used the Language Archiving Technology (ELAN) [23, 24]. Figure 2 illustrates the annotation software, with a video of a subject on the top left corner, and the coded AUs below the video. After annotation, the data was exported to an excel spread sheet as shown in Figure 3. From the data extracted, we provide a rich FACS coded database freely available for researchers in further investigation on facial behavioral analysis.

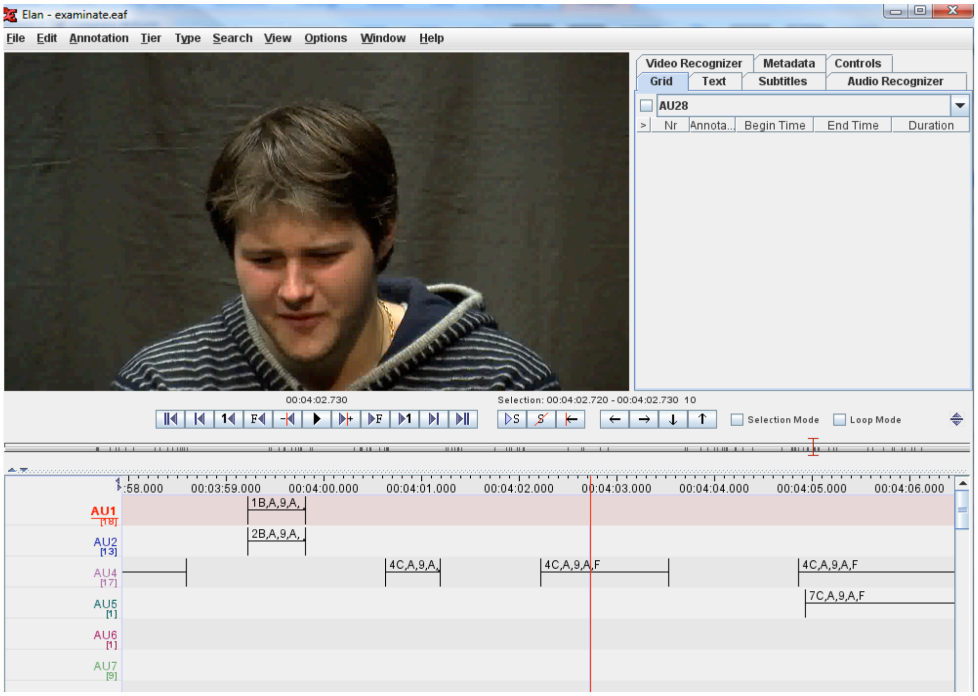

Figure 2. Illustration of the Language Archiving Technology, ELAN, used by our FACS coders in annotating the Facial Action Units.

	A	B	C	D	E	F	G	H	I
1	**Subject 02**								
2									
3	**Action Units**	**begin time**	**end time**	**duration**	**AU Intensity**	**Topic**	**Question**	**A/Q**	**GroundTruth**
4	AU1	32.89	34.28	1.39	1C	A	1	A	1
5	AU1	185.33	186.72	1.39	1C	A	8	A	1
6	AU4	70.08	71.34	1.26	4C	A	3	A	1
7	AU4	179.39	180.59	1.2	4C	A	8	A	1
8	AU12	60.61	64.36	3.75	12C	A	2	A	1
9	AU12	77.85	80.2	2.35	12B	A	3	A	1
10	AU12	93.97	96.2	2.23	12C	A	4	A	1
11	AU12	107.88	110.35	2.47	12B	A	4	A	1
12	AU12	135.3	144.4	9.1	12B	A	6	A	1
13	AU12	160.32	164.62	4.3	12C	A	7	A	1
14	AU12	210.71	212.45	1.74	12B	A	9	A	1
15	AU12	219.96	224.79	4.83	12B	A	10	A	1
16	AU12	226.15	228.95	2.8	12A	A	10	A	1
17	AU14	125.9	127.54	1.64	R14D	A	5	A	1

Figure 3.The layout of the partially exported AUs annotation into a spreadsheet.

3.3 Result Analysis and Discussion

From 32 subjects, we filtered out the subjects whom confused with the instructions and uncertain about their own intention in the interview sessions. After filtering, there were only 28 subjects available for analysis. We found 70 facial AUs in our dataset – 56 AUs from the standard FACS coding and another 14 AUs was defined to match the clues from literature review. Table 3 lists the AUs with the respective meaning. The first 56 AUs are the standard AUs in Ekman & Friesen's guidelines [18], and the last 14 AUs (Italic and bold) are our additional labeled AUs to represent other cues found in the dataset.

The unusual behavior appeared in our study included: cough, eyes move regularly to the left and right, face turned red, hand on face, quick blink, head tilt left and right, hand on neck, heavy breath, forehead muscles movement, lip pucker to the left, lip pucker to the right, scratching, quivering lips and stutter. Some of the unusual behaviors listed are relevant to culture, for instance, head tilt left and right is only observed on the group of students with Indian culture background. To further interpreting the data, we run three statistical analyses. In ***Analysis I***, we analyze the facial AUs statistically. Then we implemented machine learning methods in finding the accuracy of classification in truth tellers and liars in ***Analysis II***. Finally, ***Analysis III*** looking for the best threshold in machine learning classification – the trade-off of the cost and the risk of missing the target.

Analysis I: Statistical Analysis

We summarized the frequencies of the Facial AUs for 28 subjects, which is 280 questions and 280 answers. We examine the following research question: Were there any differences in the facial actions of the questioning states: prepare to lie and prepare to be truthful, and answering states: lying states, truthful states, telling the lies with intention of being truthful, and telling the truth with the intention of lying.

We observed that the total AUs in deceitful condition is slightly less than truthful condition. The decrease movement of liars is supported by the fact that the liars attempted control clues [21]. At a glance, we also observed that AU4, AU7, AU9, AU10, AU24, AU32, AU43, AU51, AU52, AU55, AU82, AU84, and most of the additional unusual behavior occurred more often in deceitful condition than truthful condition. These observations might be the useful cues to examine the distinction between lie and truth. For further justification, we run a statistical analysis to examine the significance of the cues.

The occurrence of AU97, AU98, AU101, AU102, and AU108 indicate lie, however, these rare events are not sufficient in monitoring the targets. For instance, a selective system to filter out the suspects based on these five cues will produce 57.14% of sensitivity, and 35.71% of false positive. It is not reliable as these AUs might also indicate anxiety. A Non-parametric test on a set of 10 AUs *{AU9, AU23, AU24, AU32, AU82, AU97, AU98, AU101, AU102, AU108}* was conducted. The primary measure used was the frequency of exhibition of the facial visual cues, it is the number of times it was exhibited. By applying the non-parametric sign test, the result for the set of 10 AUs would be considered statistically significant ($p<0.05$). This result indicated that the number of subjects who exhibited the 10 AUs is more frequent when they were in deceitful condition compare to when they were being honest.

Analysis II: Machine Learning Methods in Classification

To find if there are useful predictors of deception, we performed classifications by using machine learning experiments on the in-house dataset, the coded facial AUs. The classifications were based on 72 features: the 69 AUs (AU50 Speech is excluded), asymmetry, duration, and ground truth. Each feature represented the frequency of the AU for each question. Hence, for each participant, there will be 10x72 dimensional features for the 10 truths and 10x72 dimensional features for 10 lies. The ground truth is provided for each questions for the purpose of training, and for the machine to automatically calculate the prediction accuracy. To find out the best machine learning classifier on our in-house dataset, we used five popular classifiers implemented in WEKA software package [22], namely: Logistic Regression (LR), Multiple Layer Perceptron (MLP), Naïve Bayesian (NaiveBayes), Radial Basis Function (RBF), and Support Vector Machine (SVM). The default method in WEKA package - cross validation method with 10 folds - was experimented. Table 4 shows the comparison of the classification accuracy by using different machine learning algorithms. The best result was achieved by using a LR with sensitivity of 47.9%, specificity of 71.2%, and ROC area of 0.638. The poorest result was achieved by SVM which produced high specificity and poor sensitivity.

Table 3. The list of Facial AUs occurred in our FACS-coded database.

Type	Meaning	Type	Meaning
AU1	Inner Brow Raise	AU50	Speech
AU2	Outer Brow Raise	AU51	Head Turn Left
AU4	Brow Lowerer	AU52	Head Turn Right
AU5	Upper Lid Raiser	AU53	Head up
AU6	Cheek Raise	AU54	Head Down
AU7	Lids Tight	AU55	Head Tilt left
AU9	Nose wrinkle	AU56	Head Tilt Right
AU10	Upper lip raiser	AU57	Head Forward
AU11	Nasolabial Furrow Deepener	AU59	Head Nod
AU12	Lip Corner Puller	AU60	Head Shakes
AU13	Sharp Lip Puller	AU61	Eyes turn left
AU14	Dimpler	AU62	Eyes turn right
AU15	Corner Depressor	AU63	Eyes up
AU16	Lower Lip Depress	AU64	Eyes down
AU17	Chin Raiser	AU68	Eye Rolling
AU18	Lip Pucker	AU72	Lower Face not visible
AU19	Tongue Show	AU80	Swallow
AU20	Lip Stretch	AU82	Shoulder shrug
AU21	Neck Tightener	AU84	Head shake back and forth
AU23	Lip tightener	AU85	Head nod up and down
AU24	Lip presser	AU92	Partial Flash
AU25	Lips Part	***AU95***	***Cough***
AU26	Jaw Drop	***AU96***	***Eyes move to left & right***
AU28	Lips Suck	***AU97***	***Face turned red***
AU29	Jaw Thrust	***AU98***	***Hand on face***
AU30	Jaw sideways	***AU99***	***Quick blink***
AU31	Jaw Clencher	***AU100***	***Head tilt left and right***
AU32	Bite	***AU101***	***Hand on neck***
AU33	Blow	***AU102***	***Heavy breath***
AU36	Tongue Bulge	***AU103***	***Forehead muscles***
AU37	Lip wipe	***AU104***	***Lip pucker to the left***
AU38	Nostril Dilate	***AU105***	***Lip pucker to the right***
AU40	Sniff	***AU106***	***Scratching***
AU43	Eye Closure	***AU107***	***Quivering lips***
AU45	Blink	***AU108***	***Stutter***

Table 4. Comparison of accuracy of classification by using machine learning algorithms

Type	sensitivity	specificity	ROC
LR	47.9%	71.2%	0.638
MLP	41.0%	72.2%	0.595
NaiveBayes	42.5%	69.6%	0.580
RBF	32.6%	83.9%	0.587
SVM	31.4%	83.9%	0.595

Overall, the LR is out-performed and gives the best result in classification. But, machine learning algorithms tend to bias to truth prediction, as shown in Table 4 with low sensitivity and high specificity. This is not acceptable in real life application as it tends to miss a lot of deceptive cases and it is equal to or less accurate than by chance. Hence, we proposed a new classification threshold to increase the sensitivity, as presented in the following section.

Analysis III: Threshold of Lie and Truth

An interesting question and observation about the definition of lying in our study was: What is the percentage of lie from a subject would the session to be considered as deceptive? In a fair game, 50% threshold normally is the cutting point in decision-making for classification. We used this standard classification threshold, i.e. 50% to run an experiment. Since LR performed the best among the classifiers, it is implemented in the rest of our experiments. By cross-validate the participants with five folds (with 22 subjects as training set and 6 subjects as testing set in each fold), we achieved the result as illustrated in Table 5. Please note that the split between training set and testing set was done randomly. This produced a sequence of 30 predictions (not 28 sequence). The overall accuracy is 53.3%, with sensitivity of 36.7% and specificity of 70.0%. It was expected that we will get poor sensitivity with high specificity as the nature of machine learning algorithms favor truth prediction.

Table 5. Confusion matrix showed the accuracy by setting 50% as the threshold of lie.

Classification

Observed	Predicted		
	Lie	Truth	Percent Correct
Lie	11	19	36.7%
Truth	9	21	70.0%
Overall Percentage	55.0%	52.5%	53.3%

To overcome the bias, we made one assumption. Considering the fact that Lie is not tolerable, and we assume that if a subject lied in more than three questions in a session, then the subject is categorized as lie. This implies that we reduced the classification's threshold to a lower value, i.e. 35%. The main purpose of putting such an assumption is to reduce the false alarms and misses, as we cannot risk any possible sinister intent. The experimental result is presented in table 6, which showed the improvement of the overall accuracy to 66.7%. More importantly, it showed Increments in sensitivity to 70.0% and specificity to 63.3%.

Table 6. Confusion matrix showed the accuracy by setting 35% as the threshold of lie.

Classification

Observed	Predicted		
	Lie	Truth	Percentage Correct
Lie	21	9	70.0%
Truth	11	19	63.3%
Overall Percentage	65.6%	67.9%	66.7%

4. CONCLUSION

Problem with laboratory study of deceptive facial behavior is that it contextualized the human actions and choices [25]. It is necessary to analyze on real life data. But there is a need for cautious in putting the experimental studies into real-life application.

The challenge is how to detect deception behavior within the context of complex social interactions and how to develop paradigms in which subjects have a real choice as to whether and when to lie. The real intention of a subject to deceive the examiner is crucial. The problem of giving instruction to lie eliminates the voluntary intention to deceive. There are not consequences for the subjects' action (negatively), no harm can come to anyone and we do not achieve a valid representation of the process of deceptive acts. In the future, we have to consider the pragmatics of human communication [26] in our experimental design.

The literature review identified those psychological behaviours that might plausibly be used to detect malicious intent and deceit in the context of port immigration and customs. In particular, it addressed the behaviours that are detectable in the visual domains of facial behaviour. Our research established a rich FACS coded database that is important in future research development. In addition, in order to increase the detection rate, we proved that it is worthwhile to consider machine learning algorithms as a tool to aid human decision in human behavioural analysis.

In future work, we will investigate into multi-modalities, which combine facial behavioural analysis, body language, voice analysis, verbal content, very style and physiological methods (thermal analysis). Strong case supports from psychology research are important in spotting lies. Recently, researchers are also looking into self-deception [27]. Human is fallible in detecting deception therefore automated detection tools to augment human judgment can greatly increase detection accuracy. More research under a variety of contexts will determine which indicators and systems are the most reliable.

ACKNOWLEDGEMENT

This work was supported by EPSRC grant on "Facial Analysis for Real-Time Profiling" (EP/G004137/1). The authors would like to thank Doherty Victoria, Andy, and Ashley from QinetiQ for providing help and advices in experimental design and statistical analysis. We also express our appreciation to Bashar Rajoub, Norhayati Ahmed, Hamad Alawar, Christopher Watkins and all the volunteers in conducting the experiment.

REFERENCES

1. Bond, C.F. and B.M. DePaulo, *Accuracy of Deception Judgements.* Personality and Social Physiology Review, 2006. 10(3): p. 214-234.
2. Warren, G., E. Schertler, and P. Bull, *Detecting Deception from Emotional and Unemotional Cues.* Journal of Nonverbal Behaviour, 2009. 33: p. 59-69.
3. Bull, R., A. Vrij, and I. McKenzie, *Passenger Profiling: Assessment of verbal and non-verbal cues.* . 2002, QinetiQ report. QinetiQ/KI/CHS/CR022181.
4. Ekman, P., *Telling Lies.* Clues to Deceit in Marketplace, Politics, and Marriage. 2001, New York. London.: W.W.Norton & Company.
5. Shaw, J., S. Porter, and L. Brinke, *Catching liars: training mental health and legal professionals to detect high-stakes lies.* The Journal of Forensic Psychiatry & Psychology, 2013.
6. Collett, P., *The books of Lies: Reading People's Minds from their Actions*. 2003: London: Transworld.
7. Helman, S., et al., *Behavioural analysis to predict terrorist attacks: scoping study*. 2007, QinetiQ.
8. Ekman, P., *An argument for basic emotions.* Cognition and Emotion, 1992. 6: p. 169-200.
9. Vrij, A., *Detecting lies and deceit: The pscychology of lying and its implications for professional practice.* Chichester: John Wiley and Sons, 2000.

10. Vrij, A., *Why professionals fail to catch liars and how they can improve.* Legal and Criminological Psychology, 2004. 9(2): p. 159-181.
11. Burgoon, J.K., D.B. Buller, and L.K. Guerrero, *Interpersonal deception IX: Effects of social skill and nonverbal communication on deception success and detection accuracy.* Journal of Language and Social Psychology, 1995. 14: p. 289-311.
12. DePaulo, B.M., et al., *Cues to deception.* Psychological Bulletin, 2003. 129: p. 74-118.
13. Edwards, T. and T. Hardy, *The interactive effects of intensity and direction of cognitive and somantic anxiety and self-confidence upon performance.* Journal of Sport and Exercises Psychology, 1996. 19: p. 296-312.
14. Ekman, P., *Darwin and facial expression - a century of research in review*, ed. P. Ekman. 1973: Academic Press New York and London.
15. Russell, J.A., *The psychology of facial expressions*. 1997, Cambridge University Press.
16. Fernandez-Dols, J.M. and M.A. Ruiz-Belda, *Are Smiles a sign of happiness? Gold medal winners at the Olympic Games.* Journal of Personality and Social Psychology, 1995. 69: p. 1113-1119.
17. Kraut, R.E. and R.E. Johnston, *Social and Emotional Messages of Smiling: An Ethological Approach.* Journal of Personality and Social Psychology, 1979. 37(1539-1553).
18. Cefaratti, M. and R. Barkhi, *The Effects of Communication Media and Client Familiarity on Auditors' Confidence in Deception Detection.* Journal of Forensic & Investigative Accounting, 2013. 5(2).
19. Ekman, P., M. O'Sullivan, and W.V. Friesen, *Invited article: Face, voice, and body in detecting deceit.* Journal of Nonverbal Behaviour, 1991. 15(2): p. 125-135.
20. Buller, D.B. and J.K. Burgoon, *Interpersonal deception theory.* Communication Theory, 1996. 6(3): p. 203-242.
21. Frank, M.G., M.A. Menasco, and M. O'Sullivan, *Handbook of Science and Technology for Homeland Security*. Human Behavior and Deception Detection, ed. J.G. Voeller. Vol. 5. 2008: John Wiley & Sons, Inc.
22. Ekman, P. and W.V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, in *Consulting Psychologies Press* 1978: Palo Alto.
23. Lausberg, H. and H. Sloetjes, *Coding gestural behaviour with NEUROGES-ELAN system.* Behavior Research Methods, Instruments & Computers, 2009. 41(3): p. 841-849.
24. Sloetjes, H. and P. Wittenburg. *Annotation by category - ELAN and ISO DCR*. in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. 2008.
25. Sip, K.E., et al., *Detecting deception: the scope and limits.* Trends in Cognitive Sciences, Elsevier, 2007. 12(2): p. 48-53.
26. Verschueren, J., *Understanding Pragmatics.* Arnold, 1999.
27. Hippel, W.V. and R. Trivers, *The evolution and psychology of self-deception.* Behavioural and Brain Sciences, 2011. 34: p. 1-56.